The present invention automatically identifies speakers in an audio source by concurrently segmenting the audio source and clustering the segments corresponding to the same speaker.

Formal Objections

The disclosure was objected to because of following informalities: "homogeneous segments" is not clearly defined on page 2, line 9, and "a single full covariance Gaussian" is not clearly defined or explained on page 2, line 22.

The Examiner objected to the terms "homogeneous segments" and "a single full covariance Gaussian" as being not clearly defined or explained. Applicants submit that these terms are clearly described in the present Specification and in any event, are very well understood by a person of ordinary skill in the art. As indicated on page 2, lines 10-12, "homogeneous segments" generally correspond to the same speaker and are clustered. The manner in which "homogeneous segments" are clustered is clearly described, for example, in conjunction with the clustering subroutine shown in FIG. 4. It is clear that the segment boundaries are first identified by the segmentation routine shown in FIG. 3 to separate the various speakers. The clustering subroutine of FIG. 4 clusters the "homogeneous segments" identified by the segmentation routine of FIG. 3. See, e.g., page 7, line 21, through page 8, line 6. There is an entire section of the Specification, entitled "Speaker Clustering," beginning at page 13, line 11, that is directed to the clustering of "homogeneous segments." The criteria for when two distinct clusters are merged (i.e., corresponding to "homogeneous segments") are clearly set forth. Generally, the merge criteria is based on the very well understood BIC criterion. A reference describing BIC theory is cited on page 6, line 27.

Chen et al., "Speaker, Environment and Channel Change Detection and Cluster via the Bayesian Information Criterion," Proc. of the DARPA Broadcast News Workshop (Feb. 1998), which is incorporated by reference into the Specification on page 13, lines 20-22, describes an off-line clustering technique. The fact that this paper is authored, in part, by an inventor of the present invention does not diminish the fact that the paper describes a suitable technique for off-line clustering or is representative of what was understood by those of ordinary skill. In any event, the present specification

provides a full and complete discussion of how to cluster "homogeneous segments" that correspond to the same speaker.

The segmentation routine is based on a model selection problem. A first model, M_1, assumes no segment boundary within a window of samples (x_1, ...x_n) and is drawn from a single full covariance Gaussian. A second model, M_2, assumes a segment boundary within a window of samples (x_1, ...x_n) and is drawn from two full covariance Gaussians. See, e.g., page 8, line 20-25. The model parameters are further defined on page 9, lines 1-19. The population of each model with full covariance Gaussians is described in the specification and is well understood by anyone with a mathematics background.

Thus, Applicants respectfully request that the objection to the terms be withdrawn.

Independent Claims 1, 16, 23 and 30-35

Independent claims 1, 16, 23 and 30-35 were rejected under 35 U.S.C. §102(b) as being anticipated by Chen et al.

In the Office Action dated August 27, 2002, the Examiner asserted that Chen discloses speaker, environment and channel change detection and clustering via the Bayesian Information Criterion for segmenting the audio stream into homogeneous regions according to speaker identity, environmental condition and channel condition and clustering speech segments into homogeneous clusters according to speaker identity, environmental condition and channel (citing page 1, paragraph 2) which reads on the claimed "method of tracking a speaker in an audio source, said method comprising the steps of identifying potential segment boundaries in said audio source; and clustering homogeneous segments from said audio source substantially concurrently with said identifying step."

In the Response to Office Action dated December 26, 2002, Applicants submitted that while Chen does disclose segmenting an audio stream into homogeneous regions and clustering speech segments into homogeneous clusters, the audio stream is first segmented and then clustered. Applicants noted, as further evidence that the clustering in Chen is performed only after the audio stream has been segmented, that Section 4.1 indicates that each segment is compared to all other segments before

clustering is finalized. In addition, Section 4.2, first paragraph indicates that the data set consists of an audio file that has been "hand-segmented into 824 short segments."

In the present Office Action, the Examiner notes that the prior art cites that "our segmentation algorithm can successfully detect acoustic changes" (Chen: abstract) and that "we first examine whether our detected change points were true." (Chen: Section 3.3, paragraph 3.) The Examiner asserts that this suggests that Chen not only employs its own segmenting mechanism, but is also capable of combining segmentation with clustering "substantially concurrently."

The Examiner also asserts that Chen suggests that clustering does not need completely segmented data, such that a clustering process may be combined with a segmenting process together substantially concurrently, since Chen discloses that "it is also clear that our criterion can be applied to top-down methods." (Chen: Section 4.1, paragraph 4.)

The Examiner further asserts that a clustering step can be inserted in the segmentation loop, in Chen, Section 3.2, paragraph 1, and that Chen is capable of combining segmentation and clustering since the segmentation and clustering algorithms are based on the BIC algorithm and since equations (2), (3), and (8) have no limitation for combining segmentation and clustering.

Applicants acknowledge that Chen employs its own segmenting mechanism, but find no indication of or suggestion to perform segmentation and clustering "substantially concurrently" in the cited text. Applicants note that the Examiner asserts that Chen is capable of this, but does not assert that Chen suggests or discloses combining segmentation with clustering substantially concurrently.

Applicants note that, in the top-down method, a hypothesis is made regarding the number of clusters. Then, a test is made to determine if the number of clusters hypothesized actually "fits" the data. Alternatively, in the bottom-up method, the number of clusters is determined from the data. Thus, the capability to utilize a top-down method does not suggest that segmentation is performed substantially concurrently with the clustering process.

Regarding the final assertion made by the Examiner, Applicants also note that, whether or not Chen is capable of combining segmentation and clustering, there is no disclosure or suggestion to do so.

Thus, Chen does not disclose or suggest a "method of tracking a speaker in an audio source, said method comprising the steps of identifying potential segment boundaries in said audio source; and clustering homogeneous segments from said audio source substantially concurrently with said identifying step," as required by independent claims 1, 16, 30, 31, 32 and 33 of the present invention. Similarly, independent claims 23, 34 and 35 require that the segmentation and clustering are performed on the "same pass" through said audio source.

Additional Cited References

Kleider et al. was also cited by the Examiner in rejecting claims 15 for its disclosure that the information of the speaker model data may include a speaker name. Applicants note that the inventors listed in United States Patent Number 5,157,763 (referred to by the Examiner in the Final Ofiice Action) are not Kleider et al. Applicants did find, however, United States Patent Number 5,930,748 in the Notice of References Cited and respond to that reference below.

Applicants note that Kleider et al. is directed to a "method of identifying an individual from a predetermined set of individuals using a speech sample spoken by the individual. The speech sample comprises a plurality of spoken utterance, and each individual of the set has predetermined speaker model data." Cited, Summary of the Invention. Kleider et al. do not address the issue of segmenting speech.

Thus, Kleider et al. do not disclose or suggest a "method of tracking a speaker in an audio source, said method comprising the steps of identifying potential segment boundaries in said audio source; and clustering homogeneous segments from said audio source substantially concurrently with said identifying step," as required by independent claims 1, 16, 30, 31, 32 and 33 of the present invention. Similarly, independent claims 23, 34 and 35 require that the segmentation and clustering are performed on the "same pass" through said audio source.

Dependent Claims 2-15, 17-22 and 24-29

Dependent Claims 2 through 14, 17 through 22 and 24 through 29 were rejected under 35 U.S.C. §102 or 103 as being unpatentable over Chen, alone or in combination, with well known prior art. Claim 15 was rejected under 35 U.S.C. §103(a) as being unpatentable over Chen in view of Kleider et al.

Claims 2 through 15, 17 through 22 and 24 through 29 are dependent on Claims 1, 16 or 23, respectively, and are therefore patentably distinguished over Chen and Kleider et al., alone or in combination with well known prior art, because of their dependency from independent claims 1, 16 or 23 for the reasons set forth above, as well as other elements these claims adds in combination to their base claim.

All of the pending claims, i.e., Claims 1-35, are in condition for allowance and such favorable action is earnestly solicited.

If any outstanding issues remain, or if the Examiner has any further suggestions for expediting allowance of this application, the Examiner is invited to contact the undersigned at the telephone number indicated below.

The Examiner's attention to this matter is appreciated.

Respectfully submitted,

Kevin M. Mason

Date: June 9, 2003

Kevin M. Mason
Attorney for Applicants
Reg. No. 36,597
Ryan, Mason & Lewis, LLP
1300 Post Road, Suite 205
Fairfield, CT 06824
(203) 255-6560